Adjusting the general growth balance method for migration*

Kenneth Hill**
Bernardo Queiroz***

Death distribution methods proposed for death registration coverage by comparison with census age distributions assume no net migration. This assumption makes it problematic to apply these methods to sub-national and national populations affected by substantial net migration. In this paper, we propose and explore a two-step process in which the Growth Balance Equation is first used to estimate net migration rates, using a model of age-specific migration, and then it is used to compare the observed death rates over successive ages against residual estimates made up by the entry rate plus the net migration rate minus the growth rate. This methodology is predicated on the observation that migration rates have a very different age pattern from death rates; it is only when this condition is true that net migration and deaths can be distinguished. The procedure proposed here works reasonably well in populations that generally have good data and rather high net migration rates. There is no reason to prefer the revised migration model over the original Rogers-Castro formulation.

Keywords: General growth balance. Mortality. Migration.

Introduction

In developing countries, mortality estimates and the knowledge of the levels and trends of adult mortality are limited by the quality of available data (HILL et al., 2009a; 2009b; 2005; HILL, 2003; BANNISTER; HILL, 2004; BRASS, 1975). The most common problems faced by these countries are incomplete coverage of the vital registration system and errors in age declaration for both population and deaths counts (UNITED NATIONS, 1997; 2002).

To circumvent these data problems, different methodologies were developed to estimate adult mortality¹ and evaluate data quality. The methods used to estimate adult mortality can be classified into three major groups (TIMAEUS, 1991): distribution of deaths that evaluate completeness of death recording by comparison of the age pattern of deaths with the age pattern of the living (BRASS, 1975; BENNETT; HILL, 2004; HILL, 2003; BRASS; HILL, 1987; HILL et al., 2009a); intercensal survival (BRASS, 1975; PRESTON; BENNETT, 1983), and relative

¹ In several countries with limited databases, studies on infant mortality are already well advanced by the various methodologies available (TIMAEUS, 1991; HILL, 2003). However, there is still much doubt regarding the estimation of mortality of adult and elderly.
survival methods (BRASS, 1975; UNITED NATIONS, 1997).

The first set of methods, which we explore in this paper, considers different alternatives to estimate the degree of coverage for death counts in relation to population age structure. A common feature is that the methods estimate mortality data coverage comparing the distribution of deaths to the population age structure, using population growth rates for specific age groups (HILL, 1987). There are two major approaches: the General Growth Balance (GGB) Methods (HILL, 1987), and the Synthetic Extinct Generation (SEG) method (BENNETH; HORIUCHI, 1984). Death distribution methods make several strong assumptions: that the population is closed to migration; that the completeness of recording of deaths is constant by age; and that the completeness of population recording is constant by age, and that ages of the living and the dead are reported without error. Hill at al. (2009a) simulated several errors in age patterns in the methods and found that they perform well, although GGB is more sensitive to coverage errors, whereas BH is more sensitive to changes in census coverage. The authors find, that both methods are very sensitive to migration, but, since they both assume that migration has little or no effect on the population age structure (or growth rates), it is difficult to apply the methods to sub-national groups or countries characterized by significant migration flows.

Bennett and Horiuchi (1984), with regard to SEG and United Nations (2002), noted it is straightforward to adapt these methods for the effects of migration if the magnitude and age distribution of net migration are known. However, typically even less is known on net migration than on mortality in developing nations, and even in statistically-developed countries migration is often measured as a residual. It is not true, however, that we know nothing about migration: unidirectional migration flows, particularly those dominated by labor force movements, have age patterns that follow a broadly typical pattern that is highly concentrated in the young adult ages (ROGERS; CASTRO, 1984).

The objective of this paper is to explore the possibility of using this prior information on the age pattern of migration to fine-tune the GGB for use in a population with substantial net migration. We applied the proposed method to Mexico and Puerto Rico, both countries characterized by substantial migration flows, and compared the results from the original GGB to the methods adjusted to migration.

**General growth balance and migration**

**General growth balance method**

The General Growth Balance (GGB) method (HILL, 1987) is a generalization to all closed populations of the Brass (1975) Growth Balance method. The Demographic Balancing Equation expresses the identity that the growth rate of the population is equal to the difference between the entry rate and the exit rate. This identity holds for open-ended age segments $x^+$. The entry rate $x^+$ minus the growth rate $x^+$ provides a residual estimate of the death rate $x^+$. We can then estimate the residual from population data from two population censuses and compare it to a direct estimate using the recorded deaths (from the census or vital registration). By comparing these two records we can estimate the completeness of death recording relative to the population.

From the Balancing Equation, and considering that it holds for all open-ended age intervals, we have:

$$r(a^+) = b(a^+) - d(a^+)$$

(1)

where $r(a^+)$ is the population growth rate for the open-ended age interval, $b(a^+)$ is the birthday rate in that interval, and $d(a^+)$ is the death rate.

The “birthday” rate $x^+$ minus the growth rate $x^+$ thus provides a residual estimate of the death rate $x^+$ if the residual estimate can be calculated from population data from two population censuses and compared to a direct estimate using recorded deaths, the completeness of death recording relative to population recording can be estimated.
We can re-write equation (1) as:

\[
\frac{(N_{1+} - N_{2+})}{5} = \frac{1}{t} \ln \left( \frac{N_{2+}}{N_{1+}} \right) = \frac{1}{t} \ln \left( \frac{k_{1+}}{k_{2+}} \right) + \frac{(k_{1+} - k_{2+})}{c} \times D(x+) 
\]

in which \(N1\) and \(N2\) are population counts at two time points separated by \(t\) years, \(D\) are recorded intercensal deaths, and \(k_{1+}\), \(k_{2+}\), and \(c\) are respectively the completeness, assumed invariant by age, of the first and second population counts and intercensal deaths.

Rearranging equation (2) and, for observations affected by non-age dependent errors,

\[
b(a+) - d'(a+) = \left( \frac{1}{t} \right) \ln(k) + q_{d} \times d'(a+) \tag{3}
\]

in which \(b(a+)\) and \(d'(a+)\) are the rate of entry and exit in the population, \(r'(a+)\) is the observed growth rate, \(t\) is the intercensal interval in years, \(k\) is completeness of the first census relative to the second and \(q_{d}\) denotes the completeness of reporting deaths relative to average population coverage. The estimate of death coverage, and adjustment factor, is given by the slope of the line relating estimated death rates from population age structure and observed death rates.

The death distribution method also has an important limitation. Since it compares a distribution of deaths to an intercensal population, it estimates intercensal completeness of mortality estimates, and not the completeness at the beginning or end of the intercensal period. Thus, the age pattern of mortality reflects the age distribution of the population around the time of the second census, not the average age distribution of the population over the intercensal period. In order to take this into account, we should first use the deaths and the population from the second census to calculate age-specific death rates, and then estimate average annual deaths for the intercensal period by applying death rates to an estimate of the age distribution of the intercensal population.

Introducing migration to GGB

The assumption of both the SEG and GGB methods that the population is closed to migration is also of importance for countries that have experienced net emigration in recent decades. Both the SEG and GGB methods use information on deaths and growth rates cumulated above a series of ages \(x\). If there is some age \(x\) above which net migration is negligible, the performance of the methods above that age will be unaffected. The intercept and slope of the GGB method were obtained by orthogonal regression to points for the ages 35+ to 75+. In most cases one does not have consistent information on migration and its age schedule to define this age. Also, in order to allow cross-national comparisons it is useful to have a single method using the same assumptions. Thus, it is important and relevant to develop a method that allows for migration to be considered into calculations.

In most cases, if equation (3) is applied to a population experiencing significant positive net migration, \(k\) is likely to be overestimated and \(q\) is likely to be underestimated. Intuitively, out migrants will affect the intercensal population growth rate \(r(a+)\), at the ages in which migration is higher. Hill et al (2009a) simulated a variety of data distortions and showed that when emigration rates are high, GGBUnderestimates mortality, because out-migration affects the estimation of the intercept.

Bhat (2002) proposes a generalization of the GGB method for populations affected by migration. Equation (1) is generalized to include migration as:

\[
r(a+) = b(a+) - d(a+) + nm(a+) \tag{4}
\]

in which \(nm(a+)\) is the net migration rate for the population aged \(a\) and over. To apply Equation (4), data on intercensal migration are necessary. Bhat (2002) acknowledges that this information is not always readily available and suggests the use of a standard pattern, such as Castro-Rogers or another available one.

The method we explored here is similar to the one proposed by Bhat (2002), but introducing an iterative approach in order to obtain more robust results and allow the method to be applied in cases where migration is large and for country or sub-
populations with no information on migration patterns.

The general growth model open to migration can be written from (4) as:

\[ b(a+) - r(a+) + nm(a+) = d(a+) \]  
(5)

in which \( nm(a+) \) is the net migration. Rearranging,

\[ r(a+) + d(a+) - b(a+) = nm(a+) \]  
(6)

Let us assume that net migration rates have a typical age pattern \( \gamma m^x \). We can then rewrite (5) as:

\[ r(a+) + d(a+) - b(a+) = \frac{knm + qnm * nm^x(a+)}{1} \]  
(7)

\( qnm \) is a constant that relates the quantum of migration in the standard rate set to the quantum of migration in the actual population. \( PYL \) is the person-years lived and \( \gamma nm^x \) is a constant. The slope of the relationship in Equation (6) estimates \( qnm \); any systematic error in the population growth rate will also result in an intercept term related to the change in census coverage.

The adjusted values of \( knm + qnm * nm^x(a+) \) can now be replaced in equation (5):

\[ b(a+) - r(a+) + \frac{(knm + qnm * nm^x(a+))}{1} = \frac{(1/i)ln(k)}{1} + qd * d(a+) \]  
(9)

We expect the value \( qd \) to be close to 1. If \( qd \) turns out to be substantially different from 1.0, the first estimate from Equation (8) can be used to recalculate \( d(a+) \) in Equation (7), and re-estimate \( qnm \) until \( qd \) converges to a fixed value.

**Specifying the migration pattern**

Equation (8) requires a set of model net migration rates \( \gamma m^x \). Rogers and Castro (1984) proposed a simplified basic migration standard (based on their 11-parameter model) intended to represent a gross migration flow generated by labor force mobility. The model combines an exponentially-decaying incidence of child migration (accompanying their parents) with a double exponential to represent the fast increase and more gradual decline of worker moves. The model, however, is one of a gross migration flow. The age pattern of net migration is likely to resemble that of a gross flow only if the flow is very large in one direction – but as Rogers has written, there is no such thing as a net migrant, and therefore no such thing as a typical age pattern of net migration. However, we initially used the Rogers and Castro simplified model to provide model age-specific rates.

Using the parameter values suggested by Rogers and Castro for their simplified model, migration at age \( x \), \( m(x) \), is given by:

\[ m(x) = 0.02 * \exp(-0.1*x) + 0.1 * \exp(-0.1*(x-20)- \exp(-0.4*(x-20))) \]  
(10)

We have calculated age-specific rates for 5-year age groups by evaluating Equation (10) at single year points and averaging the rates over five years of age. Resulting rates are then assumed to represent net migration, in the hope that in settings where net migration is substantial enough to affect the GGB, it will also be dominated by a one-way flow that will be captured approximately by the Rogers-Castro model. Graph 1 shows the age-specific migration function. Graph 1 also shows an adapted Rogers-Castro model with a double exponential term for return migration centered on age 55.

**Results**

**Initial tests with Mexico 1990-2000 and Puerto Rico 1980-1990**

Mexico and Puerto Rico are two populations that have experienced substantial net emigration in recent decades, but also have reportedly complete registration of deaths. Graph 2 shows the application of the GGB without any adjustment for migration to data for males from Mwexico for the period 1990-2000 (2a) and from Puerto Rico for the period 1980-1990 (2b). The x-axis shows the observed death rate for ages \( x^+ \), and the y-axis represents the death rates for ages \( x^+ \) obtained by residual. Following Bhat (1990), we fitted an orthogonal regression to estimate a straight line to the points for the age segments considered in the GGB estimation. The slope of the fitted regression
GRAPH 1
Model age patterns of migration: Rogers-Castro simplified basic migration standard and adapted net migration model

The intercept of the line provides an estimate of the relative coverage between the two censuses used in the analysis.

model estimates the adjustment factor necessary to correct observed death rates to the death rates estimated by the residual.


GRAPH 2
Application of the GGB to data for males

a) Mexico – 1990-2000

K1/K2 = 1.018  \[ \text{Slope} = 0.913 \]

(continue)
The analysis of the graph confirms our concern about the assumption of closed populations. The points at young ages or both males and females are very irregular and off the fitted line which is consistent with problems arising from net emigration (HILL et al., 2009a). The effects of migration are clear from the kinks in the observations for younger ages and are reflected in the results that deaths appear to be over-registered relative to the population (slope < 1.0). The expectation that census coverage would be estimated to worsen from the first census to the second is only borne out for Mexico; however, in Puerto Rico, the intercept is negative, indicating an improvement in coverage from the first to the second census.

Graph 3 shows the results of applying migration estimating equations (6) and (7) to the same Mexico (3a) and Puerto Rico (3b) data. On the vertical scale is the residual estimate of the migration rate \(a^+\), and on the horizontal scale the migration rate \(a^+\) implied by the Rogers-Castro simplified model combined with the appropriate age distribution. Parameters of the straight lines fitted by orthogonal regression to the points from 15+ to 75+ are shown below the graphs. In each case, there is an apparent linear relationship, although the fit is not very good. It is also notable that the intercept is not close to zero. In the original GGB, a non-zero intercept is interpreted as indicating change in census coverage; however, in this case residual estimates of net migration for points for ages 45+ and over are positive, whereas those at younger ages are mostly negative. The interpretation of the intercept here is that it indicates return migration at older ages partially balancing the effects of out-migration at younger ages.

The next step is to insert the fitted values of \(k_{nm} + q_{nm} \times nm(a^+)\) into Equation (8) to adjust the left hand side residual estimate of the death rate \(a^+\) for the effects of migration. Results are shown in Graph 4(a) for Mexico and 4(b) for Puerto Rico. The points still have a very clear wiggle at the youngest (left-hand end) ages, but intercepts are very close to zero and slopes are very close to 1.0. There is certainly no need in either of these applications to carry through a second iteration using adjusted deaths in Equation (6). Although the migration model did not
Adjusting the general growth balance method for migration

Hill, K. and Queiroz, B.

GRAPH 3
Growth balance estimation of the intensity of migration using the simplified migration standard for males

a) Mexico – 1990-2000


Source: Data compiled in Hill (2002).

fit perfectly by any means, and apparently has not fitted net migration at the youngest ages very well, the overall fit of the GGB has improved substantially. Of particular interest
is the observation that the intercept from Equation (6) appears to account reasonably well for return migration.

The evidence in the cases of both Mexico and Puerto Rico of substantial return migration suggests that the use of a revised migration model in calculating the net migration rates \( nm(a+) \) is worth exploring. We have added an additional double exponential term to the Rogers-Castro basic simplified model to provide for return migration peaking at age 55 as shown in equation (10). Net migration rates by age implied by this revised model are shown in Graph 2.

\[
m(x) = 0.02 \times \exp(-0.1 \times x) + 0.1 \times \exp(-0.1 \times (x-20)) - \exp(-0.4 \times (x-20)) - 0.02 \times \exp(-0.125 \times (x-55)) - \exp(-0.4 \times (x-55))
\]  

Using the revised model in Equations (6) and (7) led to the results shown in Graph 5. The intercepts are closer to zero than when using the basic model, and the overall fit is somewhat better in the case of Puerto Rico, though not in the case of Mexico. Replacing the fitted values of \( k'_{nm} + q'_{nm} \times nm(a+) \) into Equation (8) to adjust the left hand side residual estimate of the death rate \( a+ \) for the effects of migration using the revised model generates the results shown in Graph 6(a) for Mexico and 6(b) for Puerto Rico. The revised model generates intercepts that are both greater than zero (better enumeration coverage of the first than of the second census) and slopes that are less than one (more complete registration of deaths than of the population).

**GRAPH 4**

Application of the growth balance equation (adjusted for simplified migration standard) to data for males


a) Mexico – 1990-2000

![Graph 4](image)

- Observed
- Fitted

\[ K1/K2 = 1.001 \]

\[ \text{Slope} = 0.998 \]

(continue)

Source: Data compiled in Hill (2002).

GRAPH 5
Growth balance estimation of the intensity of migration using the simplified net migration standard for males

a) Mexico – 1990-2000

Slope: -0.516
Intercept = 0.0028

(continued)
Adjusting the general growth balance method for migration

Hill, K. and Queiroz, B.

Source: Data compiled in Hill (2002).

GRAPH 6
Application of the growth balance equation (adjusted for the simplified net migration standard) to data for males


Slope: -0.516  Intercept = 0.0028

K1/K2 = 1.021  Slope = 0.977

(continue)
Discussion

Death distribution methods proposed for estimating death registration coverage by comparison with census age distributions assume no net migration. This assumption limits their applicability to sub-national and national populations affected by substantial net migration. In this paper, we proposed and explored a two-step process in which the Growth Balance Equation is first used to estimate net migration rates, using a model of age-specific migration, and then used to compare the observed death rates over successive ages against residual estimates made up by the entry rate plus the net migration rate minus the growth rate. This methodology is predicated on the observation that migration rates have a very different age pattern than death rates; only when this condition is true can net migration and deaths be distinguished.

We first used the Rogers-Castro (1981) Simplified Basic Migration Model to represent net migration. This is a model of gross migration rates, and when used in GGB in applications to Mexico and Puerto Rico, it is found to give intercepts that are non-zero. Inspection of the results suggests that the non-zero intercept is primarily the result of return migration in both applications. Unexpectedly (to the authors), the intercept appears to provide a reasonable way to adjust the model for return migration. Using the estimated migration rates in the GGB reaches results that indicate almost perfect death registration and consistency of census coverage in both applications.

Second, we modified the Rogers-Castro model by adding a return migration term centered on age 55. Using this model did reduce the intercepts somewhat, though it failed to improve the fits much. Using the estimated net migration rates from this model in the GGB actually produced estimates that indicated lower data quality. Although we cannot know for sure, since we do not know the truth in either application, there seems no reason to prefer the use of the modified model over the use of the original basic model. There is no reason to prefer the use of the revised migration model.
model over the original Rogers-Castro formulation.

It appears that the procedure proposed here may work reasonably well in populations that have generally good data and rather high net migration rates. Further analysis can be performed at sub-national levels in countries that experience significant internal migration. There are few studies in the literature that have applied the first version of the method (HILL; QUEIROZ, 2004) with reasonable results. Agostinho (2009) applied the method to study adult mortality in Brazilian states using the 2000 census, and Vos et al (2006) applied the method to study the burden of disease in several Asian countries. Migration schedules derived from Rogers and Castro may also be compared, when possible, to migration profiles obtained from census data or other representative survey. In any case, the method proposed here poses an advantage since it can be applied to both national and sub-national levels when migration data are not readily available.

References


Resumo

Ajustando o método da equação geral de balanceamento para migração

Os métodos de distribuição de mortalidade propostos para cobertura do registro de óbitos por comparação com distribuições de idades censitárias pressupõem população fechada ou saldo migratório pequeno. Tal pressuposto dificulta a aplicação desses métodos para populações subnacionais e nacionais afetadas por saldos migratórios substanciais. Neste artigo, propomos e exploramos um processo de dois estágios no qual a Equação Geral de Balanceamento é usada, primeiro para calcular as taxas de migração, utilizando um modelo de migração específica por idade, e então utilizado para comparar as taxas de mortalidade observadas para idades sucessivas em contraste com estimativas residuais feitas pela taxa de entrada mais a taxa de saldo migratório menos a taxa de crescimento. Esta metodologia baseia-se na observação de que as taxas de migração possuem um padrão etário muito diferente das taxas de mortalidade; apenas quando esta condição é verdadeira, saldo migratório e mortalidade podem ser diferenciados. O procedimento aqui proposto funciona razoavelmente bem em populações que possuem bons dados e taxas de saldo migratório relativamente altas. Não há motivo para preferir o modelo de migração revisto em detrimento da fórmula original de Rogers-Castro.

Resumen

Ajustando el método de ecuación general de balance para migración

Los métodos de distribución de mortalidad, propuestos para la cobertura del registro de óbitos por comparación con distribuciones de edades censitarias, presuponen población cerrada o saldo migratorio pequeño. Tal presupuesto dificulta la aplicación de estos métodos para poblaciones sub-nacionales y nacionales afectadas por saldos migratorios substanciales. En este artículo, proponemos y exploramos un proceso de dos estadios en el que se usa la ecuación general de balance, primero para calcular las tasas de migración, utilizando un modelo de migración específica por edad, y por tanto utilizado para comparar las tasas de mortalidad observadas para edades sucesivas, en contraste con estimativas residuales realizadas por tasa de entrada, más tasa de saldo migratorio, menos tasa de crecimiento. Esta metodología se basa en la observación de que las tasas de migración poseen un patrón de edad muy diferente de las tasas de mortalidad; solamente cuando esta condición es verdadera, saldo migratorio y mortalidad se pueden diferenciar. El procedimiento aquí propuesto funciona razonablemente bien en poblaciones que poseen buenos datos y tasas de saldo migratorio relativamente altas. No existe motivo para preferir el modelo de migración revisado, en detrimento de la fórmula original de Rogers-Castro.


Recebido para publicação em 30/09/2009
Aceito para publicação em 23/01/2010